Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain

AUTHOR(S):

Takeuchi, Akihide; Iida, Kei; Tsubota, Toshiaki; Hosokawa, Motoyasu; Denawa, Masatsugu; Brown, J.B.; Ninomiya, Kensuke; ... Kiyonari, Hiroshi; Ohno, Kinji; Hagiwara, Masatoshi

CITATION:

Takeuchi, Akihide ...[et al]. Loss of Sfpq Causes Long-Gene Transcriptopathy in the Brain. Cell Reports 2018, 23(5): 1326-1341

ISSUE DATE:

2018-05-01

URL:

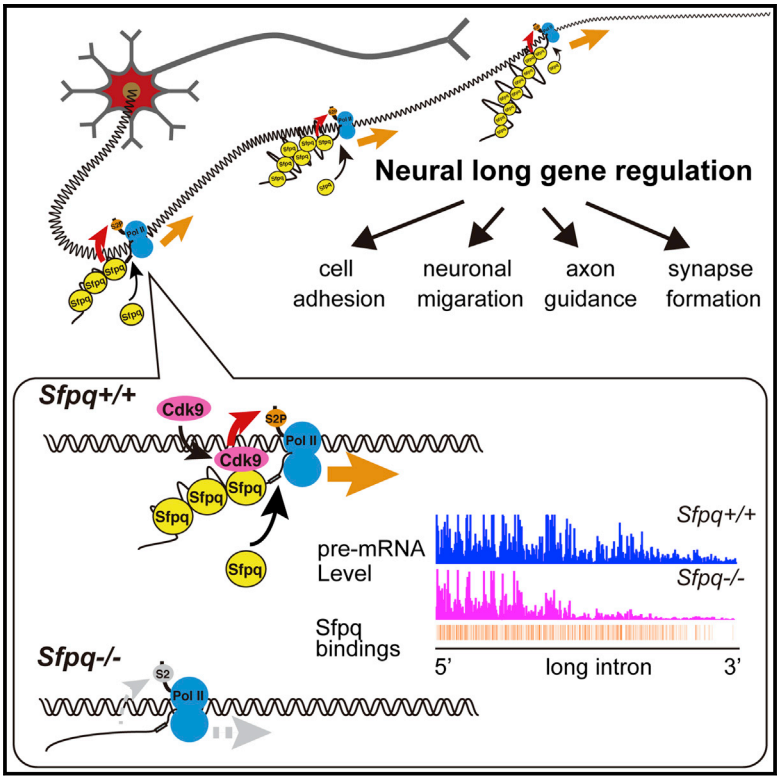
http://hdl.handle.net/2433/230954

Article

Cell Reports

Loss of *Sfpq* Causes Long-Gene Transcriptopathy in the Brain

Graphical Abstract



Highlights

- Co-transcriptional Sfpq binding to pre-mRNAs facilitates long-gene transcription
- Sfpq mediates CDK9 recruitment to the elongation complex and activates Pol II-CTD
- Sfpq selectively regulates developmentally essential neuronal genes >100 kb
- *Sfpq* disruption caused neuronal apoptosis in developing mouse brains

Authors

Akihide Takeuchi, Kei Iida, Toshiaki Tsubota, ..., Hiroshi Kiyonari, Kinji Ohno, Masatoshi Hagiwara

Correspondence

takeuchi.akihide.8r@kyoto-u.ac.jp (A.T.), hagiwara.masatoshi.8c@kyoto-u.ac.jp (M.H.)

In Brief

It has been a long-standing question how mammalian neuronal cells achieve full gene length transcription of extra-long genes. Takeuchi et al. show that RNA-binding protein Sfpq sustains long-gene transcription through Pol II-CTD activation. Loss of *Sfpq* caused long-gene transcriptopathy, which could be the cause of neurodegenerative and psychiatric disorders.

Data and Software Availability

GSE60246

Takeuchi et al., 2018, Cell Reports *23*, 1326–1341
May 1, 2018
https://doi.org/10.1016/j.celrep.2018.03.141

CellPress

Cell Reports
Article

Loss of *Sfpq* Causes Long-Gene Transcriptopathy in the Brain

Akihide Takeuchi,[1,9,*] Kei Iida,[1,2] Toshiaki Tsubota,[1] Motoyasu Hosokawa,[1] Masatsugu Denawa,[2] J.B. Brown,[3] Kensuke Ninomiya,[1,8] Mikako Ito,[4] Hiroshi Kimura,[5] Takaya Abe,[6] Hiroshi Kiyonari,[6,7] Kinji Ohno,[4] and Masatoshi Hagiwara[1,*]

[1]Department of Anatomy and Developmental Biology, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
[2]Medical Research Support Center, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
[3]Laboratory for Molecular Biosciences, Life Science Informatics Research Unit, Graduate School of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan
[4]Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan
[5]Department of Biological Sciences, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, Japan
[6]Genetic Engineering Team, RIKEN Center for Life Science Technologies, Kobe 650-0047, Japan
[7]Animal Resource Development Unit, R IKEN Center for Life Science Technologies, Kobe 650-0047, Japan
[8]Present address: Laboratory for RNA Biofunction, Institute for Genetic Medicine, Hokkaido University, Sapporo 060-0815, Japan
[9]Lead Contact
*Correspondence: takeuchi.akihide.8r@kyoto-u.ac.jp (A.T.), hagiwara.masatoshi.8c@kyoto-u.ac.jp (M.H.)
https://doi.org/10.1016/j.celrep.2018.03.141

SUMMARY

Genes specifically expressed in neurons contain members with extended long introns. Longer genes present a problem with respect to fulfilment of gene length transcription, and evidence suggests that dysregulation of long genes is a mechanism underlying neurodegenerative and psychiatric disorders. Here, we report the discovery that RNA-binding protein Sfpq is a critical factor for maintaining transcriptional elongation of long genes. We demonstrate that Sfpq co-transcriptionally binds to long introns and is required for sustaining long-gene transcription by RNA polymerase II through mediating the interaction of cyclin-dependent kinase 9 with the elongation complex. Phenotypically, *Sfpq* disruption caused neuronal apoptosis in developing mouse brains. Expression analysis of Sfpq-regulated genes revealed specific downregulation of developmentally essential neuronal genes longer than 100 kb in *Sfpq*-disrupted brains; those genes are enriched in associations with neurodegenerative and psychiatric diseases. The identified molecular machinery yields directions for targeted investigations of the association between long-gene transcriptopathy and neuronal diseases.

INTRODUCTION

From an evolutionary perspective, the pre-mRNA transcripts of vertebrates are comparatively expanded, and, in mammals, genes preferentially expressed in the brain have significantly longer introns (Gabel et al., 2015; Polymenidou et al., 2011). Given that the basic machinery of transcription is retained between lower and higher organisms but that mammalian genes expressed in the brain are significantly longer, it stands to reason that higher organisms would be more vulnerable to the dysregulation of gene processing and that some extended machinery provides augmented support for transcription in higher organisms (Oh et al., 2017). Thus far, mutations of *FUS* and *TDP-43*, two RNA-binding protein (RBP) genes regulating long intron-containing genes, have been implicated for their association with amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) (Cortese et al., 2014; Lagier-Tourenne et al., 2012; Polymenidou et al., 2011; Rogelj et al., 2012). Inhibition of topoisomerases also reduces transcription of long genes and has been linked to autism spectrum disorder (ASD) (King et al., 2013). These observations have yielded the hypothesis that some neurodegenerative and psychiatric diseases are in fact long-gene diseases or long genopathies. Yet, it has remained unclear what mechanism specifically regulates long genes to ensure their long-distance transcription.

Here we focused on neuronal RBP Sfpq (proline/glutamine rich, also known as PSF; Patton et al., 1993), which has been increasingly recognized for its roles in ALS (Thomas-Jinu et al., 2017), FTLD (Ishigaki et al., 2017), and ASD (Chang et al., 2015; O'Roak et al., 2012), suggesting its connection in some way with long genopathies. Sfpq has RNA recognition motifs (RRMs) and it plays multiple roles, such as mRNA processing, transcriptional regulation, and DNA repair (see the review Yarosh et al., 2015), though its *in vivo* functions have yet to be clarified. By generating *Sfpq*-knockout mice and performing *in vivo* and molecular analyses, we identified Sfpq as a key regulator of long neuronal gene expression and mechanisms underlying its regulatory role.

RESULTS

Sfpq Is Expressed in Newly Generated Neurons during Development

To elucidate the functions of Sfpq during development, the expression profile of Sfpq was analyzed. To assess expression

at a macroscopic level at an appropriate intermediate point in development, whole mouse embryos at embryonic day 14.5 (E14.5) were analyzed. Sfpq was robustly expressed in the whole CNS, including the spinal cord, consistent with previous descriptions (Chanas-Sacré et al., 1999; Lowery et al., 2007), and weak expression was detected in the heart, diaphragm, lung, and kidney (Figure S1A). In the cerebral cortex, Sfpq was specifically expressed on the pial surface, a site of the developing cortical plate (Figure S1A, panel Ctx). Next, a focused analysis of time-dependent *Sfpq* expression in developing mouse brains was performed (Figure 1A). Expression of *Sfpq* mRNA was evident from E12.5–E13.5, especially around the pial and apical surfaces of the developing cerebral cortex, the cortical hem (hem), and the subpallium (SP). *Sfpq* expression strengthened and increased in density on the pial surface of the cerebral cortex at E14.5. At this time, *Sfpq* expression was also observed in developing thalamus (Th) and hypothalamus (HT). *Sfpq* was detected as a robust band on the superficial region of the developing cerebral cortex by E15.5, intensified by E18.5, and was diminished by post-natal day 0 (P0). These expression patterns seemed to coincide with the distribution of newly generated neurons, including in the developing cortical plate.

To further assess *Sfpq*-expressing cells with the markers for neuronal progenitors or differentiated neurons in the cortical plate, using bromodeoxyuridine (BrdU) or TuJ1 (class III β-tubulin), respectively (Figure 1B). From E12.5 to E13.5, *Sfpq* expression was scattered in the superficial region just beneath the marginal zone, indicating its expression in nascent cortical plate neurons (CP). *Sfpq* expression was not observed in the preplate (PP, containing early born neurons) at E12.5. *Sfpq* expression was also detected in cells adjacent to the apical surface at the bottom of the ventricular zone, indicating its expression in radial glia neural progenitor cells. At E14.5, *Sfpq* expression became evident in accumulating layers of cortical plate neurons, and its expression increased until E15.5 in accordance with cortical plate development. No obvious *Sfpq* expression was detected at migrating cortical neurons in the intermediate zone (IZ) between E14.5 and E15.5. Immunohistochemical analysis using a raised antibody for Sfpq indicated mRNA expression to be consistent with protein expression (Figures 1C and S1A). These expression profiles indicated that Sfpq was specifically induced in maturating neurons that had reached the cortical plate after migration.

Sfpq Co-transcriptionally Binds across the Entire Length of Pre-mRNAs

We next asked what mRNA processing is regulated by Sfpq in maturing neurons, and thus we executed *in vivo* crosslinking and immunoprecipitation (CLIP) analysis followed by high-throughput sequencing (CLIP-seq) to identify regulatory target RNAs of Sfpq in the embryonic mouse brain. CLIP-seq studies were performed using whole embryonic mouse brain taken at E14.5. A band of endogenous Sfpq-mRNA complexes above 100 kDa was confirmed by IP sample radiolabeling and was subjected to sequencing (Figure 2A). Sfpq binding was observed along the entire length of pre-mRNA, and the binding density was high on long introns, especially in 5′ regions of Sfpq-bound genes, as exemplified by *Dcc* and *Ctnna2* (Figure 2B, CLIP tags). This binding pattern resembled the well-known sawtooth pattern, as has been observed for FUS (Lagier-Tourenne et al., 2012; Rogelj et al., 2012).

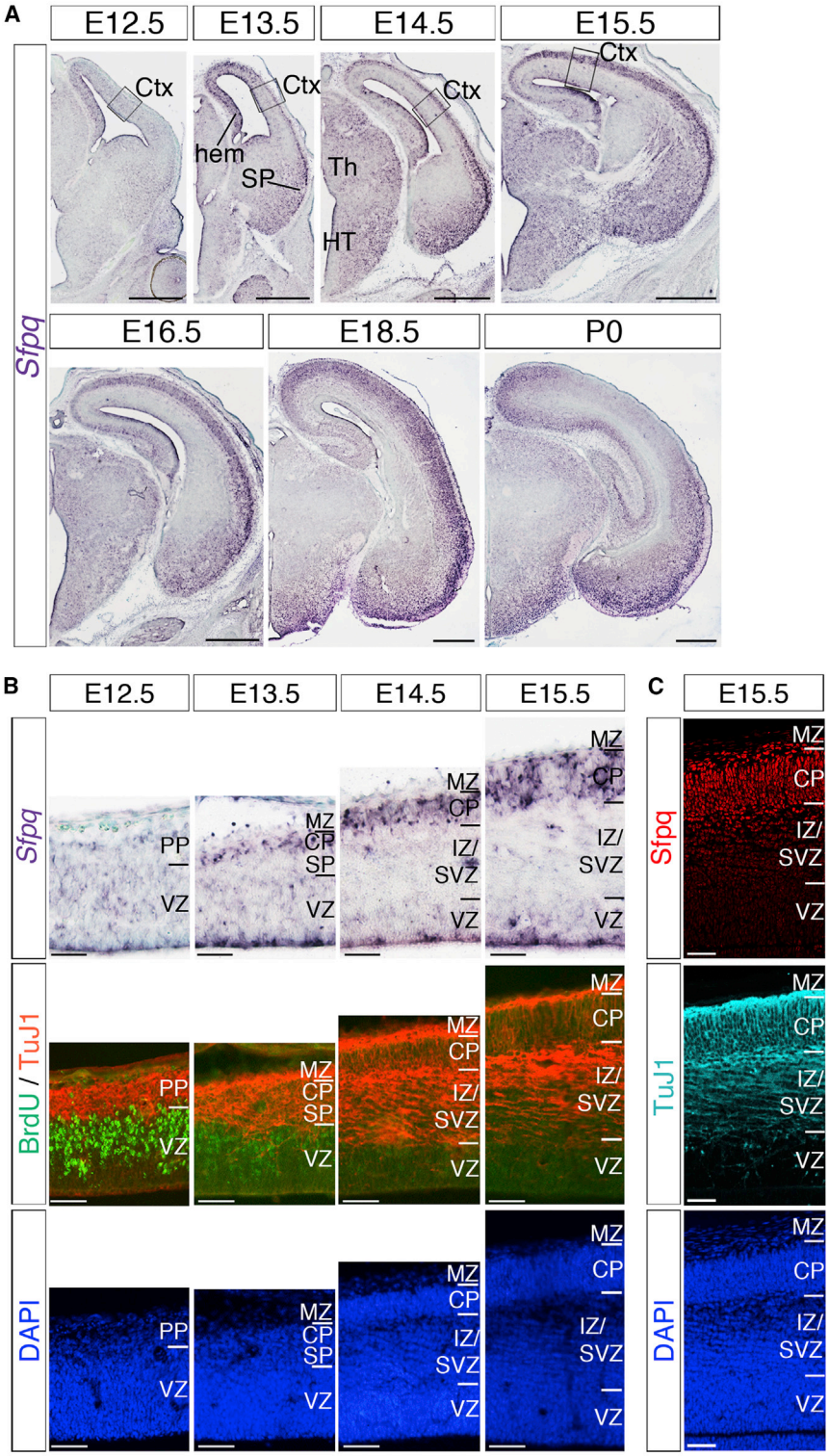
As a control for the CLIP experiment, a size-matched input control (SMInput) (Van Nostrand et al., 2016) was adopted instead of normal IgG. Sfpq-binding peaks were separated into 2 groups using binding sequence mapping p values and fold enrichment of Sfpq CLIP versus SMInput. This yielded a highly stringent Sfpq-binding peak group (High-S peaks, fold change of CLIP/SMInput [FC] ≥ 2, $p < 0.01$) and a secondary binding peak group (Low-S peaks, FC < 2, $p < 0.01$) found in both replicates. Positional distribution of Sfpq-binding peaks normalized to total expressed RNA showed that both High-S and Low-S Sfpq-binding regions were predominantly in introns (Figure 2C). Intron-binding peaks were dominantly located near the 5′ end in High-S peaks, with reductions in peaks toward the 3′ end (Figure 2D). This result is consistent with co-transcriptional recruitment of Sfpq to newly synthesized pre-mRNAs, as previously observed with FUS (Lagier-Tourenne et al., 2012; Rogelj et al., 2012).

We next tried to identify Sfpq-binding motifs. Putative Sfpq-binding motifs in mRNAs are GA-rich (Cho et al., 2014; Peng et al., 2002), GU-rich (Ray et al., 2013), AU-rich sequences (Buxadé et al., 2008) or stem structures (Peng et al., 2002; Ray et al., 2011). We examined whether broad Sfpq binding to pre-mRNAs possesses these target specificities. In our Multiple Expectation-Maximimization for Motif Elicitation (MEME) analysis, GA repeats in High-S peaks as well as GA and CA repeats in Low-S peaks were identified as significantly enriched motifs (Figure S2A), and these motifs were significantly enriched in the centers of binding peaks (Figure S2B). Next, we examined the distribution of GA and CA repeats in introns, and we found that both GA and CA repeats were homogeneously distributed along introns. Further, the frequency of these repeats was higher in long introns ($\geq$100 kb) than in short introns (<10 kb) (Figure S2C). These results indicate that Sfpq preferentially binds to long introns by utilizing non-strict-binding sequences in target pre-mRNAs.

Loss of *Sfpq* Specifically Downregulates Long Genes in the Developing Brain

Control E13.5 cerebral cortex tissues contained expression of 13,834 genes, with 11,286 (81.6%) demonstrating Sfpq binding by CLIP-seq. We developed conditional *Sfpq*-deletion mutant mice for differential evaluation (Figures S1B and S1C), and we performed *in vivo* transcriptome profiling of *Sfpq*-null (knockout [KO]) brains. The downregulation of Sfpq in KO brains was confirmed by mRNA sequencing (mRNA-seq), indicating an mRNA level of 10.5% compared to control brains, and a relative protein level of ~30% was confirmed by western blotting (WB).

A total of 192 genes were downregulated and 46 genes were upregulated in KO brains (FC ≤ 0.33 and FC ≥ 3.0, respectively; adjusted p value < 0.01, DESeq2). To systematically extract Sfpq-bound regulatory target genes, we examined the correlation between counts of Sfpq-binding peaks and gene expression changes. 152/192 (79%) of the significantly downregulated genes had High-S peaks, but no correlation was observed between counts of peaks and fold change (Figure S3A). Analysis of correlation between counts of Low-S peaks and



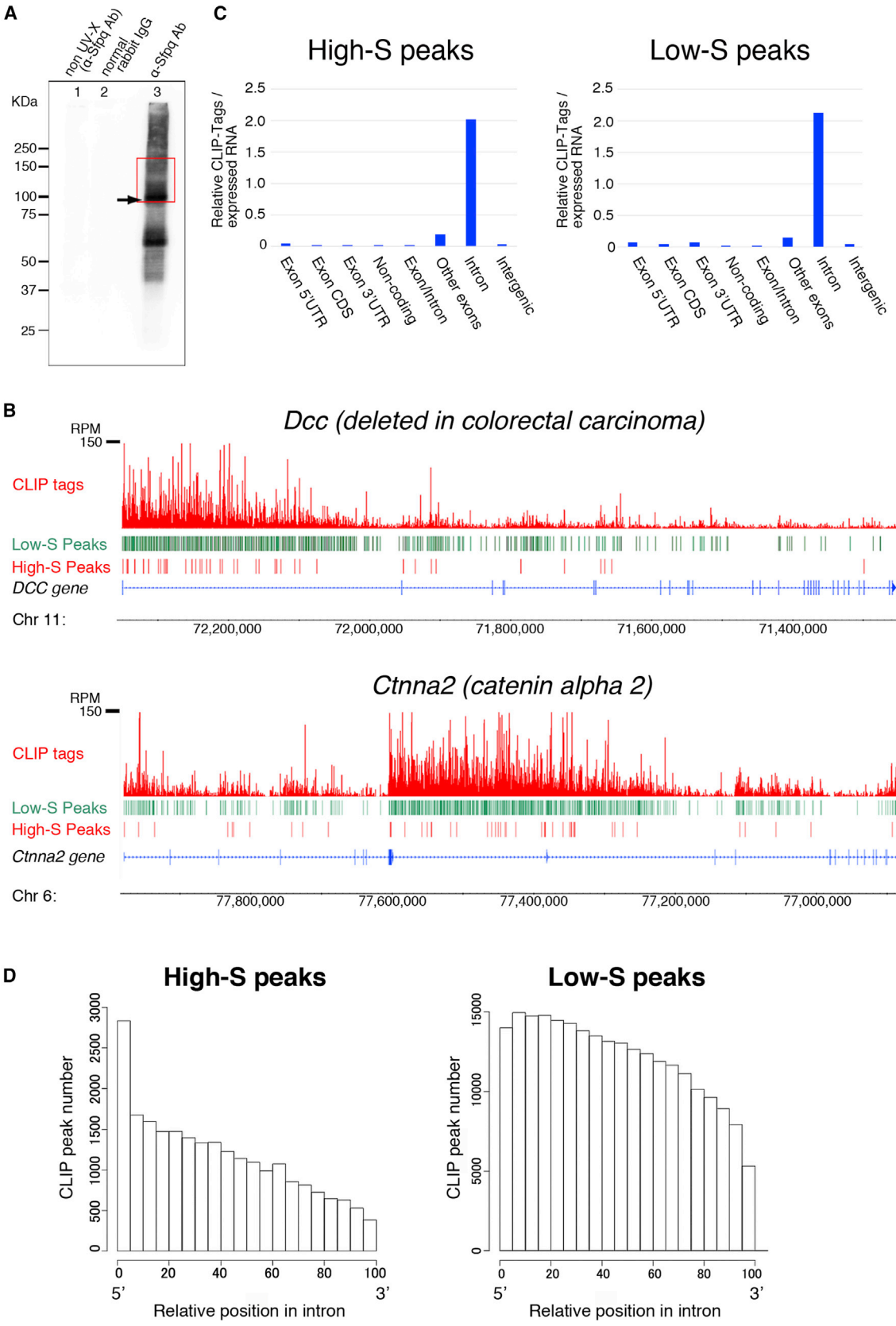
Ctx, cerebral cortex; HT, hypothalamus; SP, subpallium; Th, thalamus.
CP, nascent cortical plate; IZ, intermediate zone; MZ, marginal zone;
PP, preplate; SP, subplate; SVZ, subventricular zone; VZ, ventricular zone.

Figure 1. *Sfpq* mRNA Expression in Newly Generated Neurons during Brain Development

(A) Digoxigenin-labeled coronal sections subjected to *in situ* hybridization for *Sfpq* at E12.5, E13.5, E14.5, E15.5, E16.5, E18.5, and P0. Scale bar, 500 μm.

(B) Higher magnifications of the boxed areas in (A). Tissues were immunostained with antibodies against BrdU (green) and Tuj1 (red) and counterstained with DAPI, as shown in adjacent panel sections. Scale bar, 50 μm.

(C) Immunostaining of Sfpq (Red) and Tuj1 (turquoise) and counterstaining with DAPI at E15.5. Scale bar, 50 μm.



(legend on next page)

downregulated genes was also negative (Figure S3B). However, augmenting High-S peak presence with a threshold of 32 Low-S peak calls or more retained 93% (141/152) of High-S peak genes, and it resulted in improved enrichment of downregulated genes in KO brains (Figures S3C and S3D). The peak-FC analysis was cross-checked against pre-mRNA length. As gene length increased from 100 kb and beyond in KO brains, the FC of the qualified genes tended to be more negative (Figure 3A). This result indicated that both High-S Sfpq-binding peaks and broad Low-S peaks on entire introns, as observed in Figure 2B, were functionally significant for long pre-mRNA expression; it suggests the susceptibility of downregulation increases with gene length. Sfpq-regulatory target genes identified by FC and qualified bindings had lengths ranging from 58 kb to 1.55 Mb, with a median length of 325.5 kb; 95.7% (135/141) were longer than 100 kb (Figures S3E and S3F). We have termed this specific downregulation of extra-long genes in *Sfpq*-KO brains as long-gene transcriptopathy.

Loss of *Sfpq* Impairs Transcriptional Elongation Accompanied by a Gradual Decrease of Pol II Density

To decipher the regulatory mechanism of long-gene pre-mRNA expression by Sfpq, we examined pre-mRNA levels by employing rRNA-depleted RNA sequencing (Ribo(-) RNA-seq). Figure 3B shows the pre-mRNA expression of representative genes *Dcc* and *Ctnna2* in KO and control brains, which are each 1.1 Mb in length. To clearly detect changes in pre-mRNA expression, we determined pre-mRNA ratios between KO and control brains along the span of their lengths (Figure 3B). Interestingly, pre-mRNA expression was not downregulated in the 5′ region of pre-mRNAs for *Dcc* (ratio $\approx$1.0), and it was upregulated in the 5′ region for *Ctnna2* (ratio > 1.0) in KO brains. Expression ratios decreased along the 5′ end to the middle of transcripts, especially in long introns (ratio < 1.0), and they plateaued out from the middle of transcripts to the 3′ end.

We next examined whether the 5′-to-3′ downregulation of pre-mRNAs was observed in all genes yielding High-S peaks in the Sfpq CLIP-seq analysis. Genes were assessed for pre-mRNA levels along the span of their lengths. For genes <100 kb, there was little change of upregulation in the 5′ region, and downregulation in the 3′ region of pre-mRNAs was observed in KO brains compared to controls (Figure 4A, 1,538 genes). In contrast, for genes $\geq$100 kb, relative pre-mRNA levels in KO brains were slightly increased for the 5′ region, as exemplified by *Ctnna2*, but they were significantly reduced after 100 kb (Figure 4A, 1,004 genes). Longer genes were divided into regions at regular intervals, and the fraction of reduced expression was shown to increase as position tended to the 3′ end (Figure 4A, upper right panel). These data indicated that the loss of *Sfpq* widely impaired the expression of Sfpq-binding genes and caused a gradual decrease of pre-mRNAs, which became evident after 100 kb, therefore suggesting that long genes are more susceptible to *Sfpq* loss.

We next examined whether the impact on Sfpq binding *in vivo* was similarly observed *in culture* using gene silencing with small interfering RNA (siRNA). CLIP and transcriptome studies were analogously performed in differentiated Neuro2a cells to mimic *in vivo* maturating neuronal cells. Knockdown (*Sfpq*-KD) efficiency in Neuro2a cells was confirmed by mRNA-seq, indicating an mRNA level of 8.4% compared to control cells, and by WB, indicating protein KD to ~30%. We observed identical binding patterns as well as effects of *Sfpq* disruption on long genes *in culture* (Figure S4A). A total of 176 qualified (vide infra) *in culture* genes was downregulated, including an overlap of 40 genes with *in vivo* samples (Figure S4B); 38 of the 40 overlaps were genes longer than 100 kb, and the remaining two overlaps were more than 95 kb in length. The concordance of the *in vivo* and *in culture* analyses suggests the critical functionality of *Sfpq* in either condition, and the consequence of *Sfpq* KO could not be attributed to developmental failure or off-target effects in *Sfpq*-KO brains.

Gradual decreases of pre-mRNA levels suggested the possibility of transcriptional elongation impairment by the loss of *Sfpq*. Thus, we examined the distribution of RNA polymerase II (Pol II) using chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq). As *in vivo* brain samples are not amenable for ChIP-seq, *in culture* analyses using *Sfpq*-KD Neuro2a were carried out. 176 qualified (vide infra) *in culture* genes showed gradual decreases in pre-mRNA levels (Figure S5A). We assessed Pol II distribution along the span of their lengths using a heatmap of Pol II density for significantly downregulated genes in *Sfpq*-KD Neuro2a having qualified Sfpq binding in CLIP (Figure 4B). Relative Pol II density of KD/cont Neuro2a cells was slightly increased for the 5′ region, especially on the transcriptional start site (TSS), but it was significantly decreased toward the poly(A) site (pA).

To see whether downregulation of Pol II density is directly responsible for the decrease of pre-mRNA levels, we evaluated the relationship between FCs of pre-mRNA levels and those of Pol II densities across the entire length of Sfpq target genes. We observed a positive correlation of change in Pol II densities and RNA-seq levels of those upregulated within the TSS to 100 kb and downregulated thereafter (Figures S5C and S5D). To see the change of Pol II density more clearly, we drew Pol II metagene plots of KD and control Neuro2a cells (Figure 4C). We observed a promoter-proximal pausing peak downstream of the TSS in both KD and control, indicating that transcriptional

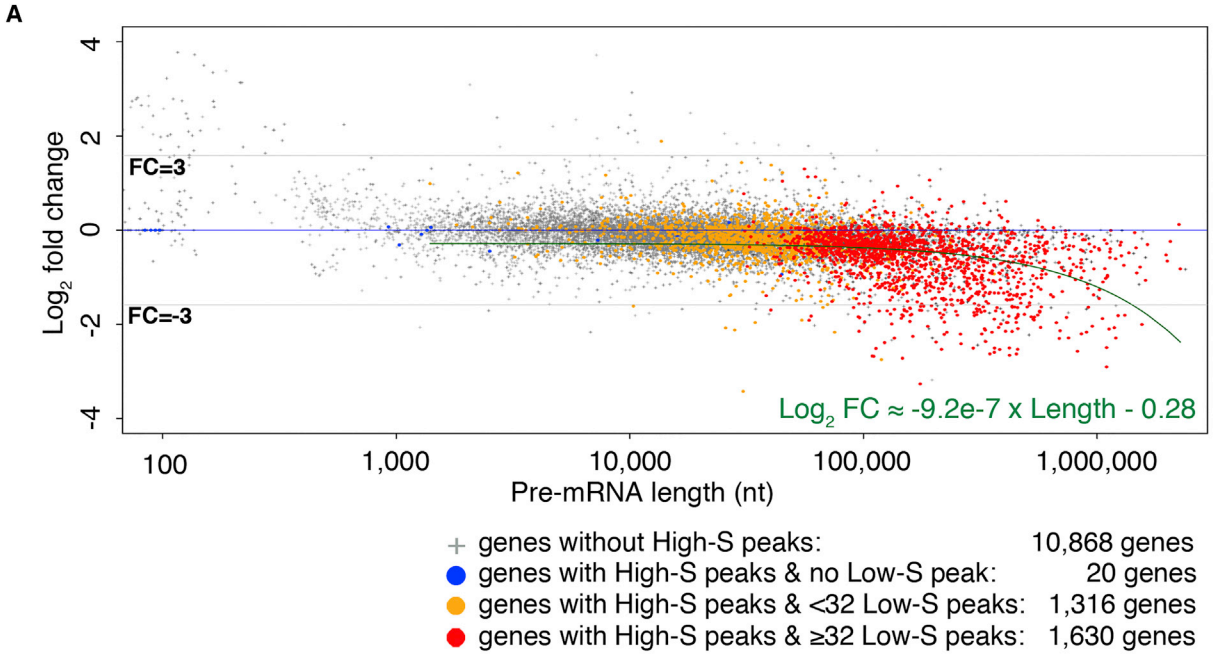
Figure 2. Neuronal RNA-Binding Protein Sfpq Is Co-transcriptionally Bound to Nascent Pre-mRNAs in Developing Neurons

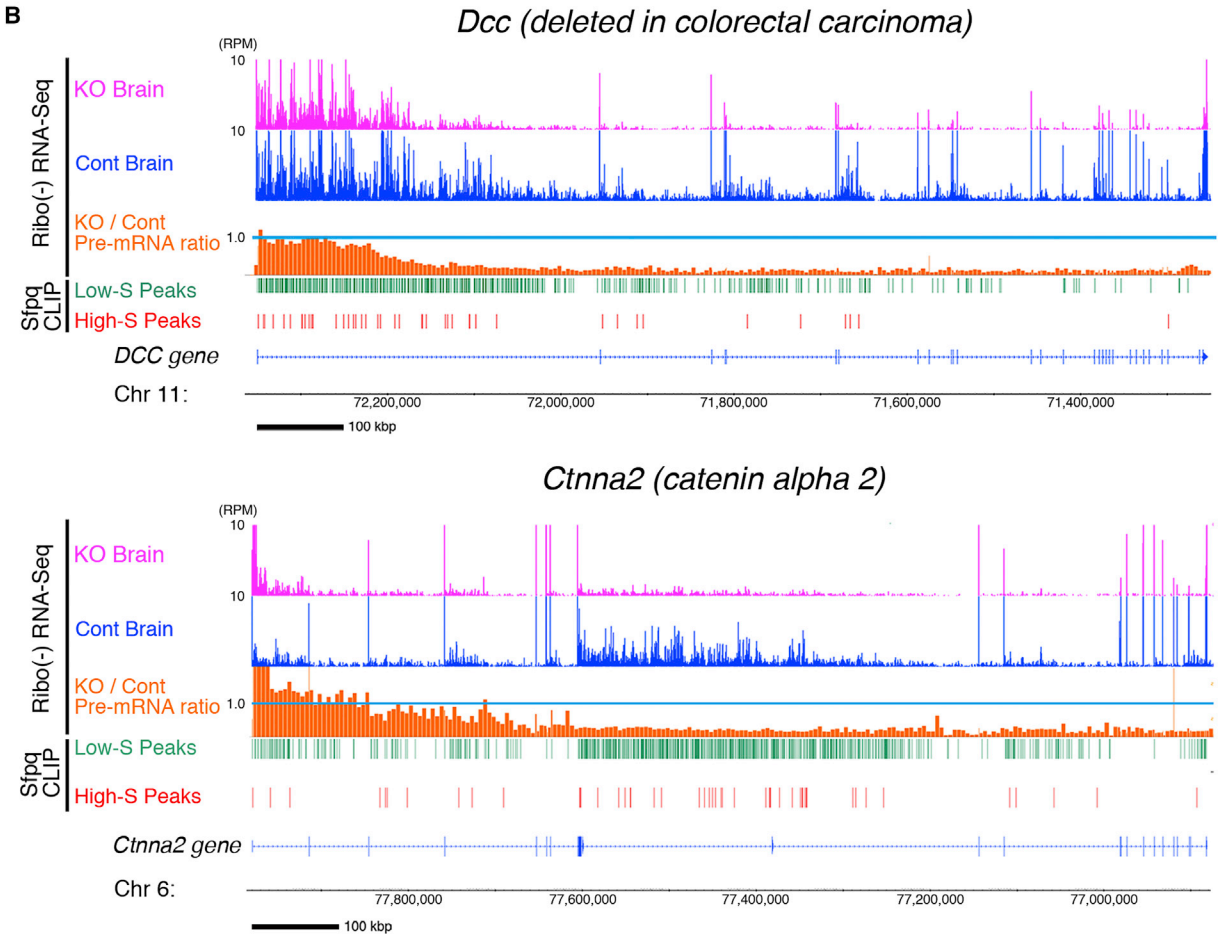
(A) CLIP samples radiolabeled at the 5′ ends were gel-electrophoresed and transferred to a nitrocellulose membrane (lanes 1–3): lane 1 shows autoradiography of Sfpq IP samples using non-UV-crosslinked input (non UV-X, α-Sfpq Ab); lane 2 shows the IP sample with normal IgG using UV-crosslinked input (normal rabbit Ab); and lane 3 shows Sfpq IP samples with UV-crosslinked input (α-Sfpq Ab). The arrow indicates the position of free protein around 100 kDa. The region 100–175 kDa (highlighted by a red box) represents shifted Sfpq-RNA complexes subjected to library construction, analyzed in tandem with a corresponding size-matched input (SMInput) control region.

(B) Distribution of Sfpq CLIP-seq tags and positions of High-S and Low-S peaks on representative genes *Dcc* and *Ctnna2*.

(C) Distribution of Sfpq High-S and Low-S CLIP peak positions, normalized against total expression in Ribo(−) RNA.

(D) Histogram showing the relative positions of Sfpq High-S and Low-S CLIP peaks in introns.





(legend on next page)

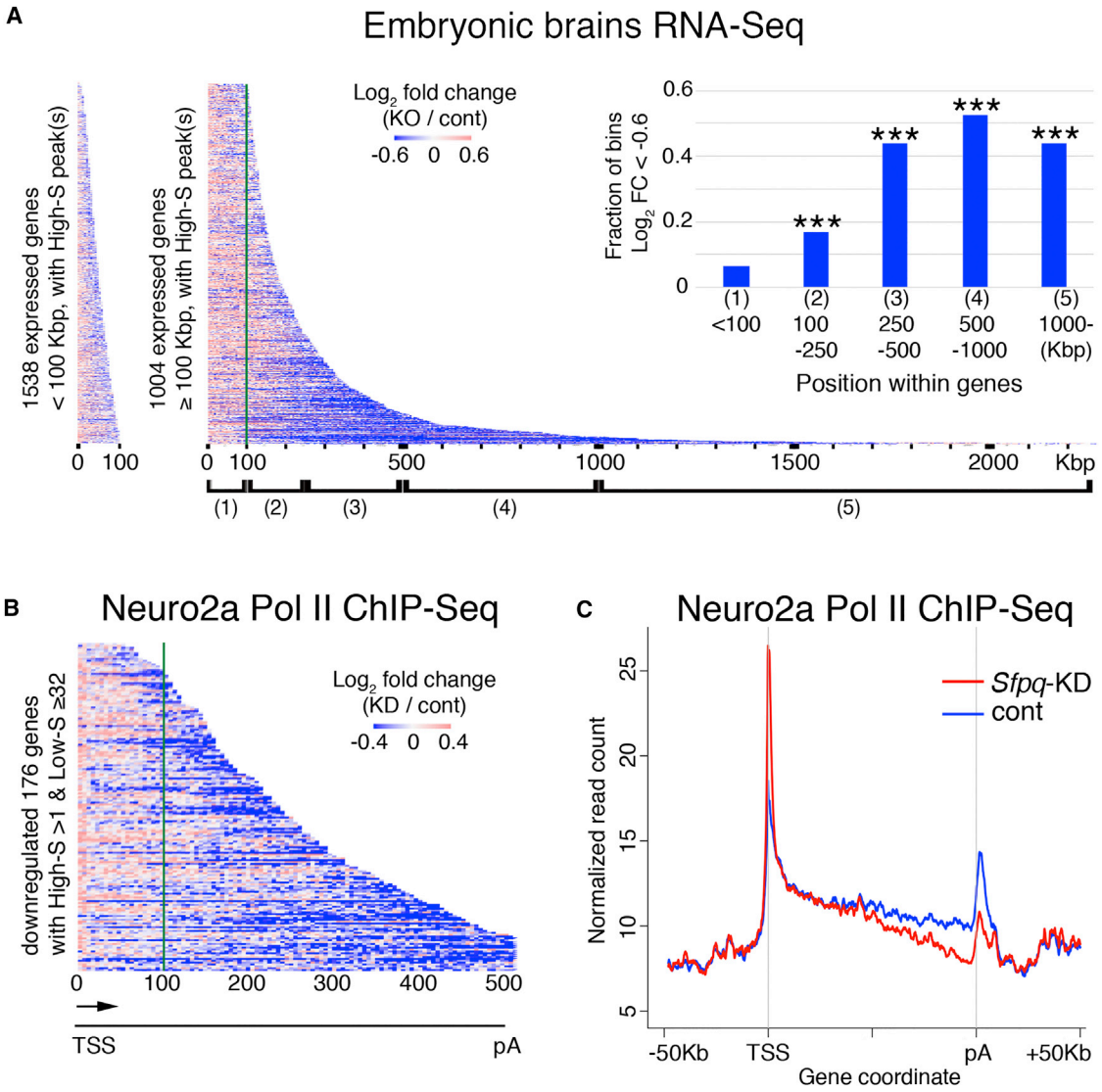


Figure 4. Loss of *Sfpq* Impaired Transcriptional Elongation with a Gradual Decrease of Pol II Density in Gene Body

(A) Heatmap of relative pre-mRNA levels in KO versus control brains for genes with High-S peaks. Fold changes are computed with a window size of 5 kb per bin. (Upper right) The fraction of bins in which the Log_2 fold change was <-0.6, corresponding to a reduction in raw expression by at least one-third, is shown (***$p < 0.001$, Fisher's exact test).

(B) Heatmap of relative Pol II density in KD versus control Neuro2a for downregulated 176 genes with qualified Sfpq bindings (High-S $\geq$ 1 and Low-S $\geq$ 32). Fold changes are computed with a window size of 5 kb per bin.

(C) Metagene analysis of Pol II ChIP-seq reads for the same gene sets as (B), relative to the TSS and pA with their $-$50-kb upstream and +50-kb downstream regions, respectively.

initiation was not impaired in KD, and the Pol II ratio at the TSS and downstream was rather higher in KD than control. Observed Pol II accumulation at the TSS and downstream may be due to stacking of Pol II from the impairment of elongation, which was often observed in the elongation defect; the compensation by transcriptional activation for downregulated gene expression; or de-repression of Sfpq, given that Sfpq can act as a transcriptional repressor through direct promoter binding (Iacobazzi et al.,

Figure 3. Loss of *Sfpq* Specifically Downregulated Genes Longer Than 100 kb

(A) Relationship between pre-mRNA length, fold change in mature mRNA, and Sfpq-binding peaks for 13,846 genes that were expressed in KO and control brains. The green line is a linear regression of genes with High-S peak(s), calculated in an x-linear, y-log space. The slope of the fitted line is captioned on the plot.

(B) *Sfpq* disruption attenuates pre-mRNAs from the 5′ end. The distribution of Ribo(−) RNA-seq tags, KO/Cont pre-mRNA ratio, and positions of Sfpq High-S and Low-S CLIP-seq peaks on representative genes *Dcc* and *Ctnna2* are shown (KO and control cortices).

2005; Imamura et al., 2014; Song et al., 2004; Urban et al., 2000). To see whether gradual downregulation of Pol II and pre-mRNA is specific for Sfpq-bound pre-mRNAs, we additionally observed the change of pre-mRNA levels and Pol II densities in 90 downregulated genes (FC $\leq$ 0.33), which did not have qualified Sfpq bindings. In many of these genes, downregulation of Pol II density was frequently observed within 100-kb regions from the TSS and remained low to pA, suggesting promoter-dependent downregulation (Figure S5B). These results indicated that *Sfpq*-KD caused impairment of transcriptional elongation on target genes, which would cause a gradual decrease of pre-mRNAs, that selectively affected the expression of genes longer than 100 kb.

Disruption of *Sfpq* Interrupted the Transcriptional Elongation of Long Genes by Decreased Phosphorylation of the Pol II CTD and Defective Recruitment of CDK9

In transcription, phosphorylation of Pol II on serine 2 (Ser2) and/or serine 5 (Ser5) residues in its C-terminal domain (CTD) is essential for activation (Marshall et al., 1996; O'Brien et al., 1994; Saunders et al., 2006), where Ser5 is phosphorylated during initiation and Ser2 is subsequently phosphorylated during productive elongation (see reviews Egloff and Murphy, 2008 and Odawara et al., 2011). ChIP followed by qPCR (ChIP-qPCR) was performed using phospho-specific antibodies, as we had found that total Pol II density in long genes expressed in Neuro2a was not sufficiently high for analysis using ChIP-seq alone. We utilized *Cadm1* and *Atrnl1* as representative long genes (>100 kb) because the expression of *Dcc* and *Ctnna2* was not sufficiently observable *in culture*. Prior to ChIP-seq, analogous CLIP tag analyses in wild-type as well as expression change in *in culture* KD experiments confirmed quite similar results (Figure S4C). In Pol II ChIP-qPCR using a pan Pol II antibody, which recognizes both phospho- and non-phospho-Pol II, total Pol II occupancy of *Cadm1* and *Atrnl1* was greater at the TSS than in the gene body for controls (*si-Cont*), indicating promoter-proximal pausing. In *Sfpq* KD (*si-Sfpq*), total TSS Pol II accumulation was greater than in controls as we observed in ChIP-seq (Figures 4B and 4C), but levels were decreased in the gene body, especially in the 3′ region, suggesting the impairment of transcriptional elongation (Figure 5A, Pol II). In Pol II ChIP-qPCR using phospho-specific antibodies, *Sfpq* disruption decreased the abundance of Ser2P (Figure 5A, Ser2P), but it had no effect on Ser5P (Figure 5A, Ser5P), indicating that productive elongation required Sfpq while initiation was not affected by *Sfpq* disruption. These results suggested that impairment of Ser2 phosphorylation arrested Pol II elongation, resulting in a gradual decrease in Pol II distribution across entire gene bodies (Figure 5A, Pol II, also observed in Figures 4B and 4C) and subsequent downregulation of long-gene pre-mRNA levels in a 5′-to-3′ fashion.

To further analyze the role of Sfpq in Ser2 phosphorylation of the Pol II CTD, we examined interactions of Sfpq with Pol II, CDK9, and CDK12, as both CDK9 and CDK12 activate Pol II through phosphorylation of Ser2 in the Pol II CTD (Bartkowiak et al., 2010; Marshall et al., 1996; O'Brien et al., 1994; Saunders et al., 2006). We introduced exogenous FLAG-Sfpq or FLAG-EGFP into Neuro2a cells, and we performed anti-FLAG immunoprecipitation (IP). Anti-Sfpq antibody precipitated FLAG-Sfpq with endogenous Sfpq indicating Sfpq-Sfpq interaction (Figure 5B, αSfpq, lane 4**)**. Precipitated endogenous Sfpq was not diminished by RNaseA treatment (Figure 5B, αSfpq, lane 5), indicating that Sfpq interacted with itself in an RNA-independent manner. We similarly detected coIP of Pol II (Figure 5B, α Pol II, lane 4**)** as previously described (Emili et al., 2002; Kameoka et al., 2004; Rosonina et al., 2005). In this context, coIP of CDK9 was also observed (Figure 5B, αCDK9, lane 4), consistent with a previous study in which the Sfpq interactome identified included CDK9 (Yang et al., 2015). We observed higher levels of precipitated CDK9 than Pol II, suggesting a stronger association of Sfpq-CDK9 than Sfpq-Pol II in the elongation complex. As the Pol II CTD Ser2 is phosphorylated by the P-TEFb complex (a complex of CDK9 and Cyclin T1), we examined coIP of Cyclin T1 (Figure 5B, αCyclinT1, lane 4), but the level of precipitated Cyclin T1 was much lower than that of CDK9, supporting the close association of Sfpq with CDK9 in the elongation complex. CoIP of CDK12 was examined but a relevant band in coIP samples could not be detected, indicating the selective binding of Sfpq (Figure 5B, αCDK12, lane 4**)**. We also confirmed the interaction among endogenous Sfpq, CDK9, and Pol II by IP using anti-Sfpq antibody (Figure 5C). We detected the coIP of Pol II and CDK9 (Figure 5C, αPol II and αCDK9, lane 4**),** and we found that the level of precipitated CDK9 was higher than Pol II, as we had observed in FLAG-Sfpq IP (Figure 5B).

The observed Sfpq-CDK9 interaction and the impairment of transcriptional elongation in *Sfpq*-disrupted cells strongly indicated that Sfpq mediated the interaction of Pol II and CDK9. To test this hypothesis, we immunoprecipitated protein from *Sfpq*-KD Neuro2a cells with an anti-Pol II antibody (Figure 5D). In control samples (*si-Cont*), IP of Pol II precipitated Sfpq and CDK9. In KD conditions (si-*Sfpq*), we observed reduced Ser2P-Pol II in input (Figure 5D, αS2P, lane 2), and the amount of CDK9 precipitated with Pol II was significantly reduced (Figure 5D, αCDK9, lane 4), indicating Sfpq-dependent interaction of CDK9 with Pol II. These data present us a model to explain that Sfpq co-transcriptionally bound on nascent pre-mRNAs facilitates transcription by serially recruiting CDK9 to the elongation complex and sustaining the transcriptional elongation of long-intron-containing genes (summarized in Figure 5E). As for the mechanism of co-transcriptional binding, a recent crystallography study indicated that Sfpq binds to a high-affinity site on nucleic acids and then non-specifically binds on target nucleic acids as a multimer via its coiled-coil interaction motif (Lee et al., 2015). The Sfpq CLIP showed that prominent High-S-binding peaks in 5′ introns could represent high-affinity-binding sites that facilitate subsequent multimerization toward the 3′ region, leading to broad Low-S peaks (Figures 2B and 2D). This model plausibly explains why stringently defined binding peaks (High-S peaks) alone did not distinguish target pre-mRNAs from non-target pre-mRNAs.

Sfpq Regulates Long Neuronal Genes Associated with Neurodegenerative and Psychiatric Disorders

Analyses to characterize the *in vivo* phenotypes of *Sfpq* were undertaken by developing conditional *Sfpq*-deletion mutant mice (Figures S1B and S1C). Initially, we found that homozygous

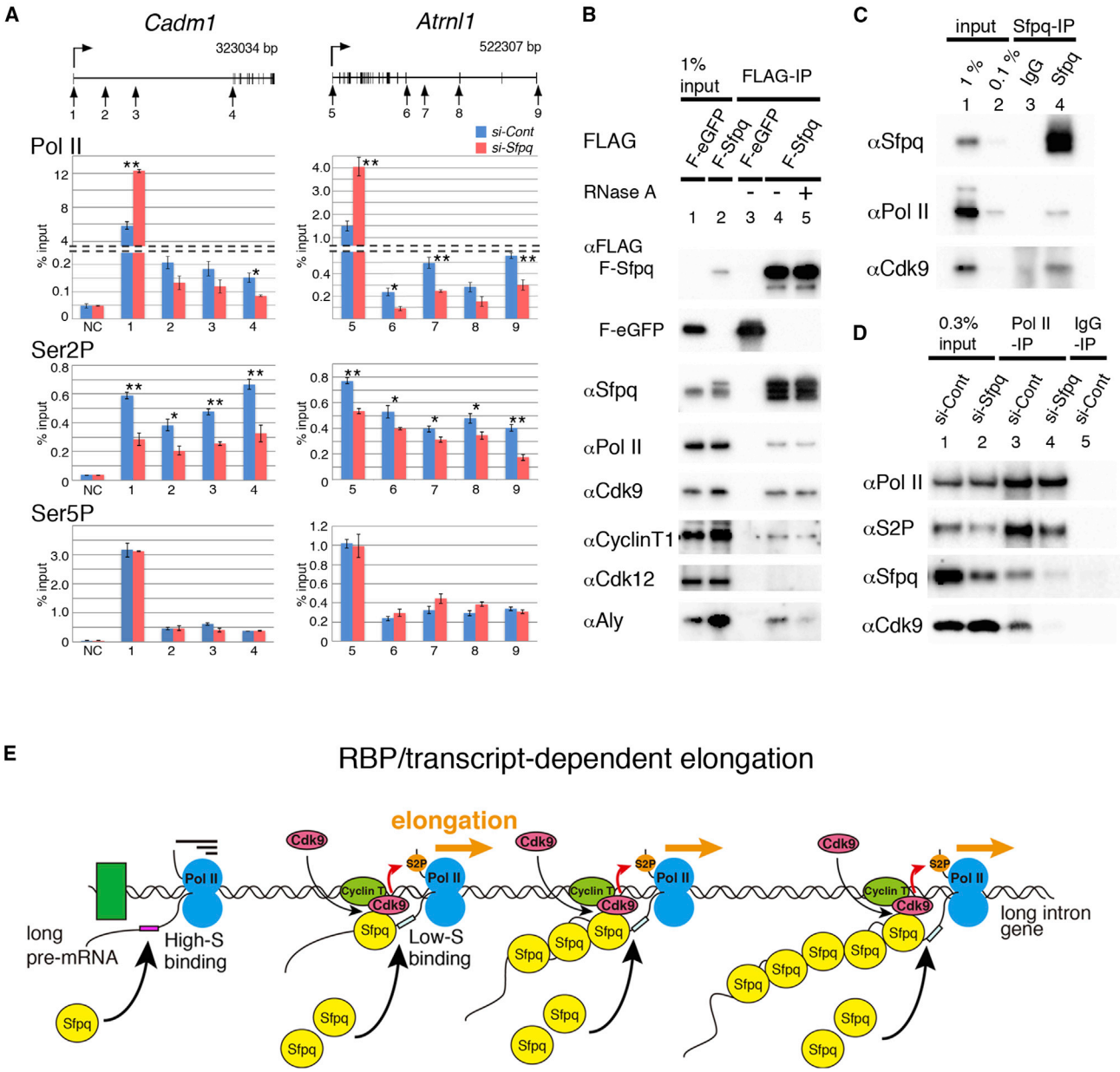


Figure 5. Disruption of *Sfpq* Interrupts the Transcriptional Elongation of Long Genes by Decreased Phosphorylation of the Pol II CTD and Defective Recruitment of CDK9

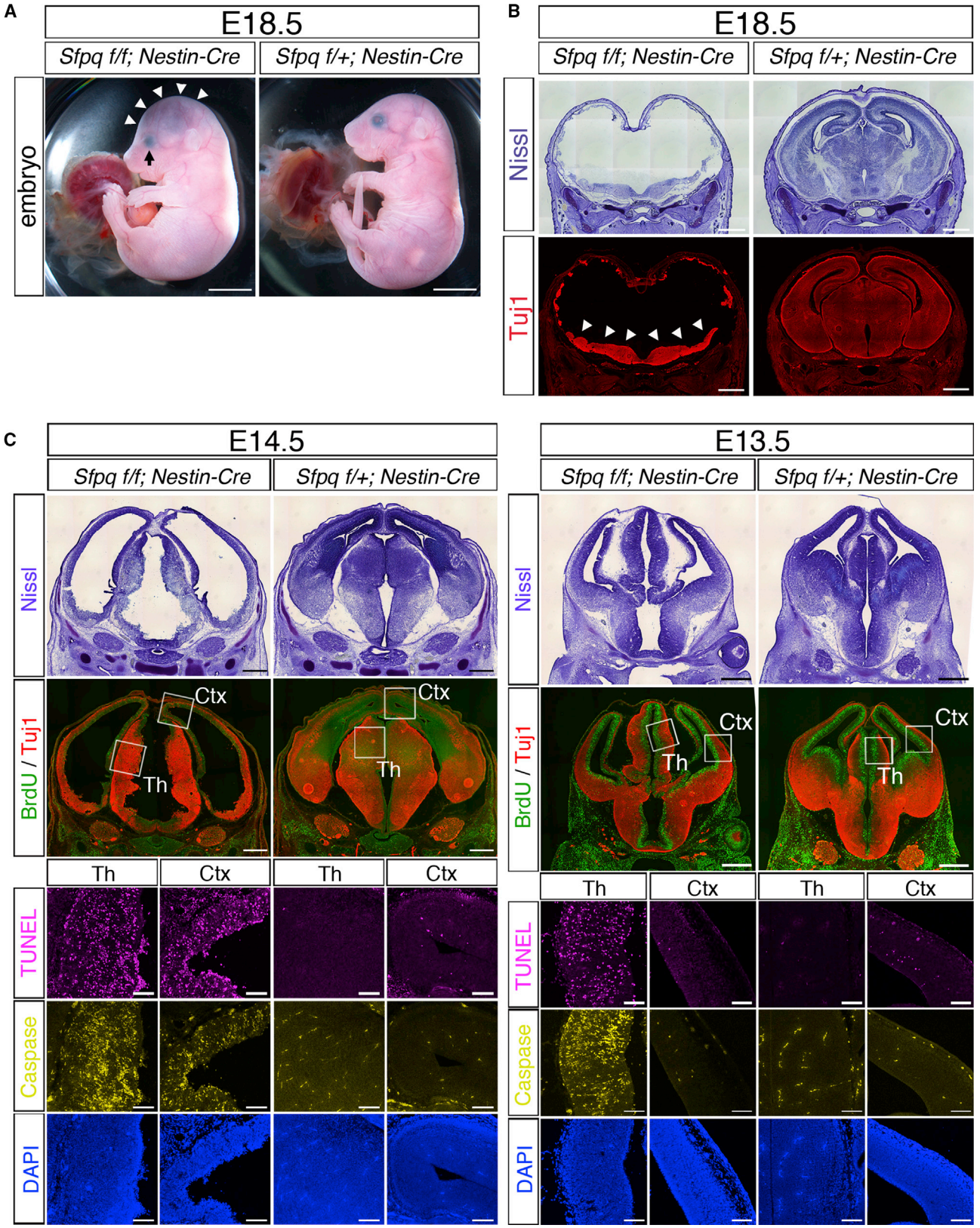
(A) ChIP-qPCR analysis of Pol II, Ser2, and Ser5 in *Cadm1* and *Atrnl1* genes. NC indicates the negative control region. Data are presented as the mean ± SEM (n = 3; *$p < 0.05$ and **$p < 0.01$ using Student's t test). Note the break in the vertical axis for Pol II.

(B) Co-immunoprecipitation (coIP) of FLAG-tagged GFP (F-GFP, lanes 1 and 3) or FLAG-tagged Sfpq (F-Sfpq, lanes 2, 4, and 5). Input samples are in lanes 1 and 2 and FLAG-IP samples are in lanes 3–5. Lane 5 shows samples treated with RNaseA before IP.

(C) CoIP of endogenous Sfpq and their associations with Pol II and CDK9. Input samples are in lanes 1 and 2 and Sfpq-IP samples are in lanes 3 and 4.

(D) CoIP of Pol II in control (si-Cont, lanes 1, 3, and 5) and *Sfpq* knockdown (si-Sfpq) (lanes 2 and 4). Input samples are in lanes 1 and 2 and Pol II-IP samples are in lanes 3–5.

(E) Proposed model of RBP/transcript-dependent elongation. High-S Sfpq bindings are formed in 5′ introns, presumably using high-affinity sites, and subsequent multimerizations make broad Sfpq Low-S bindings toward 3′ introns. This co-transcriptional binding of Sfpq to pre-mRNAs acts as recruiters of CDK9 to repeatedly facilitate Pol II elongation of long genes. S2P, Ser2P of Pol II CTD.



(legend on next page)

Sfpq$^{-/-}$ mutants died before E9.5, indicating early embryonic lethality that was likely due to the crucial role of Sfpq in development (Figure S6A). To investigate the function of Sfpq in nervous system development, we crossed *Sfpq*-floxed mice with *Nestin-Cre* mice. Although we could not obtain viable homozygous mutant post-natal mice, however, embryos were obtainable up through E18.5 (Figure S6B). Homozygous mutant embryos of E18.5 showed almost normal appearances and body sizes; however, unusual dark red structures were visible inside their heads (Figure 6A, white arrowheads) and their eyes were smaller (Figure 6A, black arrow). Histological analysis of homozygous mutant brains showed that most dorsal parts, including the cerebral cortex, were lost, while the medioventral region positive for Tuj1 immunostaining remained (Figure 6B, white arrowheads). These results indicated that Sfpq is essential for brain formation, especially in the dorsal region, including the cerebral cortex. We next analyzed brains from E11.5 to E14.5 to study the onset of developmental brain disorders using homozygous *Sfpq f/f;Nestin-Cre* mutant embryos (KO brain) and heterozygous *Sfpq f/+;Nestin-Cre* mutant embryos as a control.

Beginning with the longest-developed brains, we identified gross abnormalities in KO brains at E14.5 (Figure 6C, E14.5); in coronal sections of E14.5 embryonic brains, the thicknesses of both the cerebral cortex and the thalamus were severely reduced in KO brains compared to controls. In a histological analysis, large numbers of TUNEL-positive cells were observed in the thalamus and developing cerebrocortical regions of KO brains. Cleaved caspase-3 immunostaining was also positive in these regions, indicating that the loss of *Sfpq* induced apoptosis in the developing brain by E14.5 (Figure 6C, E14.5 TUNEL/Caspase). E13.5 embryonic brains were next analyzed to compare phenotypical consequence as well as to investigate the molecular pathology underlying the reduced thickness observed in KO brains (Figure 6C, E13.5). In coronal sections, structural differences between KO and control brains, including in the shapes of the cerebral cortex and thalamus (Th), were observed by Nissl staining. We noted a wider third ventricle in KO brains and structural mismatch when compared with controls at E13.5. TUNEL and cleaved caspase-3 immunopositive cells were clearly observed in the thalamus, but not in the cortical region of KO brains, indicating that the loss of *Sfpq* caused cell death in the thalamus prior to that in the cerebral cortex (Figure 6C, E13.5 TUNEL/Caspase). We observed TUNEL- and cleaved caspase-3-positive cells in BrdU-positive regions. Because *Sfpq* expression was detected in cells adjacent to the apical surface, it is suggested that *Sfpq* was also expressed in radial glial cells, and loss of *Sfpq* induced cell death in neuronal progenitors.

To further characterize molecular abnormalities in the KO brain, we studied the proliferation of neuronal progenitor cells in E13.5 cortices before detectable signs of apoptosis. The proliferation of neuronal progenitor cells was analyzed using BrdU-pulse labeling. No significant differences were observed with the number of BrdU-positive dividing cells in the cortical region (Figure S6C), indicating that *Sfpq* had no significant effects on neuronal stem cell proliferation. We next analyzed the expression profiles of molecular markers in E13.5 using mRNA-seq (Figure S6D). Neural stem cell markers were expressed at similar levels in KO and control cortices. No significant differences were observed in the expression levels of neural progenitor markers (*Sox2*, *Nestin*, and *Mash1*), cortical arealization markers (*Pax6* and *Emx2*), or glial cell markers (*Hes1* and *Gfap*). With respect to cortical layer markers, expression levels were relatively unchanged, although a slight reduction in the expression of the upper layer markers *Brn2* and *Cux1* was observed alongside a slight increase in the expression of the lower layer marker *Tbr1* in KO cortices. Taken together with the results of the transcriptome study performed on E13.5 cerebral cortices and the fact that abnormalities were not present in subsequent analyses of E12.5 and E11.5 brains, evidence suggests that the loss of *Sfpq* did not significantly impair the proliferation of neuronal stem cells or their neuronal differentiation but specifically downregulated genes longer than 100 kb and triggered apoptosis after E13.5.

Finally, neuron developmental processes regulated by Sfpq were investigated. We classified 141 Sfpq-regulated gene groups based on their major functions, and we identified groups essential for cell adhesion (n = 17 genes), axonal guidance (n = 15), synaptic proteins or ion channels (n = 17), and membrane proteins (including receptors, transporters, and glycoproteins) (n = 33), in addition to transcription factors (n = 20), RBPs (n = 8), and others (Figure 7A; Table S1). Additionally, gene ontology analyses of the 141 Sfpq-regulated genes concurred potential regulons (functional clusters of genes) highly enriched in nervous system development, specifically in axonal guidance, and also in dendrites and neuronal projections (Figure 7B). These annotations indicate an essential role of Sfpq in gene expression during neurophysiological development, suggesting a relationship between Sfpq dysfunction and neurodegenerative and neuropsychiatric disorders.

Genetic studies of ASD and schizophrenia (Geschwind, 2011; Veltman and Brunner, 2012) have revealed that their candidate genes are exceptionally long and encode synaptic scaffolding proteins, receptors, and cell adhesion molecules (Bourgeron, 2015). Other studies of ASD putatively assigned Sfpq as a causative gene (Chang et al., 2015; O'Roak et al., 2012). Given these

Figure 6. *Sfpq* KO Causes Apoptosis in Maturating Neurons

(A) E18.5 *Sfpq f/f; Nestin-Cre* and *Sfpq f/+; Nestin-Cre* whole embryos. "*f(Δneo)*" is designated as "*f*." White arrowheads indicate abnormal brain tissue transparently visible from the surface, and the black arrow points to an example of the small eyes observed in *Sfpq f/f; Nestin-Cre* embryos. Scale bar, 5.0 mm.

(B) Coronal sections of head regions from E18.5 *Sfpq f/f; Nestin-Cre* and *Sfpq f/+; Nestin-Cre* embryos. Nissl staining of sections (upper panels) and adjacent section immunostained for Tuj1 (lower panels) are shown. White arrowheads indicate the remaining medioventral regions positive for Tuj1 immunostaining. Scale bar, 1.0 mm.

(C) Coronal sections of the cranial regions of E14.5 embryos (*Sfpq f/f Nestin-Cre* and *Sfpq f/+;Nestin-Cre*) and analogous E13.5 embryos. Nissl staining and adjacent sections immunostained for BrdU (green) and Tuj1 (red) are shown. Scale bar, 500 μm. Boxed areas indicated in BrdU/Tuj1 are shown with higher magnification in respective TUNEL and Caspase stainings. For TUNEL, Caspase, and DAPI, scale bar represents 100 μm. Ctx, cerebral cortex; Th, thalamus.

A

GO Term	Gene #	SFARI Gene	SZ Gene
Cell adhesion	17	4	4
Axon guidance	15	5	
Synaptic protein, Ion Channel	17	6	2
Membrane protein (receptor, transporter, glycoprotein)	33	6	6
DNA-binding protein	20	3	1
RNA-binding protein	8	2	
Metal-binding protein	6		
Others	18	1	1
Unknown function	7		
total	141	27	14

B

	GO ID	GO Term	p-value
Biological Process	GO:0007411	axon guidance	1.34E-09
	GO:0007399	nervous system development	8.12E-05
	GO:0000122	negative regulation of transcription from RNA polymerase II promoter	0.01277
	GO:0045892	negative regulation of transcription, DNA-templated	0.27507
Cellular Component	GO:0030425	dendrite	0.00049
	GO:0043005	neuron projection	0.00177
	GO:0048471	perinuclear region of cytoplasm	0.14238
	GO:0000139	Golgi membrane	0.16780
Biological Function	GO:0003700	transcription factor activity, sequence-specific DNA binding	0.10097

Figure 7. Sfpq Regulates Developmentally Essential Genes Associated with Neurodevelopmental and Neuropsychiatric Disorders

(A) The data show the 141 Sfpq regulatory target genes organized by function such that each gene was assigned to a unique ontology. Each ontology was assessed for overlap with genes ascribed to ASD/schizophrenia.

(B) The data show the significant process, component, and function ontologies associated with the same set of qualified genes.

findings, we further compared the 141 Sfpq-regulated genes to the high-confidence causative gene list for autism and schizophrenia using the Simons Foundation and Autism Research Initiative (SFARI) database (https://gene.sfari.org) (Banerjee-Basu and Packer, 2010) and the SchizophreniaGene (SZGene) database of the Schizophrenia Research Forum (http://www.szgene.org/) (Allen et al., 2008); 27 and 14 genes were included in the SFARI and SZGene databases, respectively (Figure 7A; Table S1), and their enrichment was significantly higher in SFARI ($p = 1.41E{-}09$) and SZGene ($p = 4.14E{-}03$) relative to all expressed genes. Moreover, Sfpq-regulated genes have been associated with ALS, spinocerebellar ataxia, epilepsy, Alzheimer's disease, bipolar disorder, intellectual disability, and some brain tumors (MalaCards; Rappaport et al., 2017) (Table S1).

DISCUSSION

In transcription, Ser5 and Ser2 of Pol II CTD are sequentially phosphorylated during productive elongation (see the reviews Egloff and Murphy, 2008 and Odawara et al., 2011). So far, how CDKs, activators of Pol II CTD by phosphorylation, are recruited locally to elongation complexes has remained unclear. Accordingly, we found that Sfpq is required for the recruitment of CDK9 to phosphorylate Ser2 of the Pol II CTD (schematic in Figure 5E). This model can clearly explain why the loss of Sfpq decreased the Pol II density on gene bodies in a 5′-to-3′ fashion and arrested the transcription of pre-mRNAs from long genes. Requirement of RBPs for transcriptional regulation had been demonstrated such that SRSF2 bound on promoter-associated nascent mRNAs is essential for the Pol II transition from promoter-proximal pausing to the elongation phase through an interaction with P-TEFb and the phosphorylation of Ser2 of the Pol II CTD (Ji et al., 2013; Lin et al., 2008). We found that Sfpq facilitates transcriptional elongation of long genes using a mechanism similar to SRSF2. The SRSF2 studies and the work herein advances mechanistic insight into the molecular mechanisms of pausing release and subsequent activation of Pol II CTD for elongation by RBPs, respectively. The RBPs FUS and TDP-43 have also been shown to co-transcriptionally bind to long genes (Cortese et al., 2014), and their disruption causes the downregulation of long genes in the adult brain (Lagier-Tourenne et al., 2012; Polymenidou et al., 2011). Our data support the conclusion of these previous studies that RBPs can be involved in transcriptional regulation in a gene type- or cell type-specific manner.

We demonstrated co-transcriptional binding of Sfpq to nascent pre-mRNAs and Sfpq-Sfpq interaction, concordant with a study showing self-multimerization of Sfpq in gene regulation (Lee et al., 2015). These results suggest the possibility that Sfpq multimers act like histones for RNA in long introns, presumably with other intron-binding RBPs, and facilitate transcriptional elongation through a Sfpq-CDK9-Pol II interaction, as well as activate mRNA processing and stabilize pre-mRNA. So far, Sfpq has been shown to directly bind to the Pol II CTD and facilitate the splicing (Emili et al., 2002; Kameoka et al., 2004; Rosonina et al., 2005), 3′ end processing (Lutz et al., 1998; O'Connor et al., 1997), transcriptional termination (Kaneko et al., 2007), and nuclear retention (Chen and Carmichael, 2009), suggesting the function of Sfpq in co-transcriptional mRNA processing (Bentley, 2014; Hsin and Manley, 2012). These observations have brought about the consensus that the Pol II CTD acts as a recruiter of several RBPs, resulting in the coupling of transcription and pre-mRNA processing (Fong and Zhou, 2001; Muñoz et al., 2010; Singh and Cooper, 2012), as well as subsequent enhancement of transcriptional elongation with appropriate termination (Zhou et al., 2012). However, it is not well known whether RNA processing machineries also influence transcription. In this study, we demonstrated that Sfpq on nascent pre-mRNAs could act as a recruiter of CDK9 to an elongation complex and sustain transcriptional elongation, suggesting that RBPs have a more central role as coordinators of transcription and pre-mRNA processing in contrast to existing proposed peripheral roles of RBPs only in post-transcriptional mRNA regulation.

Sfpq was initially cloned as a splicing factor; we analyzed alternative splicing changes in KO brains, where we found limited splicing change (Table S2), and we also observed that Sfpq

binding was not significantly enriched around the alternative exons in KO brains (Table S3). Considering that Sfpq-regulated genes contain several RBPs, it is unlikely that *Sfpq* disruption directly caused alternative splicing changes observed in KO embryonic brains, but there still remains a possibility that these splicing changes affected the *in vivo* phenotype. A recent study showed that Sfpq directly binds to the 3′ UTR of mRNAs and modulates microRNA (miRNA) targeting of mRNAs with long 3′ UTRs (Bottini et al., 2017). Thus, we analyzed our RNA-seq data with a focus on the 3′ UTR bindings of Sfpq. However, in KO brains, we could not find a clear relationship with specific upregulation of genes that have Sfpq bindings to their 3′ UTR (Figure S7A), nor was there an obvious connection between 3′ UTR length and expression change in genes for those with Sfpq bindings (Figure S7B). Our observation indicates that a major function of Sfpq in developing brains is the regulation of transcriptional elongation.

A recent study has identified that co-transcriptional spliceosome assembly by Prp5 could act as a transcriptional elongation checkpoint (Chathoth et al., 2014). In addition, inhibition of splicing by spliceostatin A (SSA) for U2 small nuclear ribonucleoprotein (snRNP) or by an antisense oligonucleotide targeting U2 snRNA decreased the Ser2 phosphorylation of Pol II in the CTD, and it caused gene-specific 3′ end downregulation of pre-mRNAs in genes longer than 15 kb in length (Koga et al., 2014, 2015). From these observations, splicing factors and spliceosome formation are an essential activating component of Pol II via Ser2 phosphorylation during transcriptional elongation of average-sized genes that also could act as the checkpoint for normal transcription and co-transcriptional splicing regulation. Thus, it indicated that extended introns in long genes require a splicing-independent mechanism for sustaining their transcription. It has been shown that the length of the first intron in eukaryotic genes is significantly longer than all downstream introns within a gene (Bradnam and Korf, 2008). Therefore, it can be suggested that transcriptional elongation tends to stall within their 5′ regions beyond 100 kb for long genes possessing an extended first intron when *Sfpq* is disrupted. In support of this idea, the downregulation of pre-mRNAs after *Sfpq* disruption was observed beyond the 100-kb intra-intron position in typical long genes (*Dcc*, *Ctnna2* in Figure 3B, and *Cadm1* in Figure S4C); when genes were exceptionally proceeded by multiple short introns in their 5′ regions, prior to extended introns, disruption was even further downstream, as observed for *Atrnl1* (Figure S4C).

It has been proposed that RBPs comprehensively regulate functional clusters of genes referred to as regulons through specific bindings to target mRNAs by utilizing RNA-binding domains and consensus sequences (Cosker et al., 2016; Keene, 2007). In the case of Sfpq, 6.1% (135/2,197) of expressed long genes more than 100 kb were specifically regulated by Sfpq, and regulatory targets are highly enriched in genes essential for neuronal development. These data indicate that Sfpq does not aberrantly sustain all long genes but rather specifically regulates genes essential for neuronal development as regulons. In addition to RBPs, a recent study showed that U1 snRNP suppresses premature cleavage and polyadenylation and is selectively required for sustaining long-distance transcriptional elongation of extended introns in large genes, which has been termed telescripting (Oh et al., 2017). They proposed that the relative amount of U1 could be the transcription elongation checkpoint criterion, in which sufficient U1 with other factors allows long-gene expression in more tissue-specific and differentiated cells, such as differentiating neurons. These observations indicate that long genes require specific regulation for their full gene length transcription, and it is interesting that several RNA-binding molecules of RBPs/snRNP regulate different gene sets through their binding specificities.

Genetics studies have identified hundreds of risk genes associated with autism and schizophrenia (Berg and Geschwind, 2012; Geschwind, 2011; Iossifov et al., 2012; Levy et al., 2011; Sanders et al., 2012; Veltman and Brunner, 2012), which encode synaptic scaffolding proteins, receptors, and cell adhesion molecules (Bourgeron, 2015), and the candidate genes for many are exceptionally long. Importantly, many of these genes were found to overlap with Sfpq target genes in our study. Considering databases that catalog genes associated with neuropsychiatric diseases, we found that the average gene size in SFARI is 217.3 kb (King et al., 2013) and in SZGene is 112.0 kb (Sun et al., 2009) (respectively, 46.5% and 23.7% of listed genes are longer than 100 kb), whereas the average gene size of all protein-coding genes in build GRCh38.81 of the human genome is 62.7 kb, with 83% of genes shorter than 100 kb. This suggests that the uncommonness of long genes may make them susceptible to impairment, which ultimately leads to disease states. In fact, genetics studies have identified mutations of Sfpq in a large collection of *de novo* mutations associated with ASDs (Chang et al., 2015; O'Roak et al., 2012). In addition, nuclear depletion of Sfpq was observed in neurons and astrocytes in brain areas affected by Alzheimer's disease and Pick's disease (Ke et al., 2012). Thus, the dysfunction or dysregulation of Sfpq could underlie a variety of neurological diseases and psychiatric disorders. Additional analyses are required to further delineate interacting molecules and mechanisms for long-gene transcriptopathy, including a focused study on the localization and interaction among Sfpq, CDK9, and phospho-specific Pol II, which will lead to the enrichment of our understanding of the global structure of transcription machinery and the molecular mechanisms and pathological consequences of long-gene transcriptopathies.

EXPERIMENTAL PROCEDURES

Mice

Conditional KO mice for *Sfpq* were generated (Figures S1B and S1C). Detailed information is given in the Supplemental Experimental Procedures. All animal care and experiments were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals, and all experimental protocols were approved by the Institutional Animal Care and Use Committee of the Kyoto University Graduate School of Medicine and the RIKEN Kobe Branch.

CLIP-Seq

CLIP was performed as described elsewhere (Licatalosi et al., 2008; Ule et al., 2003; Van Nostrand et al., 2016), with some modifications. An anti-Sfpq antibody and a normal IgG control antibody were used. CLIP-seq libraries were generated directly from isolated RNAs using the Ion Total RNA-Seq Kit (Life Technologies). High-throughput sequencing (HTS) was performed using the Ion Proton System. Detailed information is given in the Supplemental Experimental Procedures.

Chromatin Immunoprecipitation of RNA Pol II, S2P, and S5P for ChIP-qPCR

For ChIP using Pol II antibodies CMA601, CMA602, and CMA603 (Stasevich et al., 2014), nuclear fraction was extracted using a truChIP High Cell Chromatin Sharing Kit with SDS Shearing Buffer (Covaris, MA, USA), and it was sonicated using a Covaris S220. Immunoprecipitation was performed as described previously (Blecher-Gonen et al., 2013; Kimura et al., 2008). ChIP samples were analyzed by qPCR. Primers use in qPCR and detailed information are given in the Supplemental Experimental Procedures.

Bioinformatic Analysis for CLIP

Peak calling was performed according to the enhanced CLIP (eCLIP) method (Van Nostrand et al., 2016). After peak calling, we checked the results from "Peak_input_normalization_wrapper.pl" included in eCLIP, and we defined peaks for which p values < 0.01 and the fold change was above SMInput ≥ 2 in both Sfpq IP Rep-1 and Sfpq IP Rep-2 (IP-1 and IP-2) as highly stringent (High-S) peaks. Detailed information is given in the Supplemental Experimental Procedures.

Identification of Consensus Sequences for Sfpq-Binding Sites

We employed MEME to identify the consensus sequences of Sfpq-binding sites in the target pre-mRNAs (Bailey et al., 2009). Detailed information is given in the Supplemental Experimental Procedures.

DATA AND SOFTWARE AVAILABILITY

The accession number for all data reported in this paper is GEO: GSE60246. Lists of HTS data and related information are provided in the Supplemental Experimental Procedures. For calculating reads per kilo base per million mapped reads (RPKM) and transcript per million (TPM) and for drawing plots, in-house scripts were used. All data and scripts not included here are available from the corresponding author upon reasonable request.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and three tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.03.141.

ACKNOWLEDGMENTS

We would like to thank the Kyoto University Medical Research Support Center, its Institute of Laboratory Animals, and its Radioisotope Research Center for their technical support. We also thank A. Hagiwara, A. Fujishita, M. Nakagawa, K. Wanezaki, K. Kusumoto, A. Utsumi, and Y. Watanabe for technical assistance. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT, JSPS KAKENHI 19500269, 25500288, 21249013, 15H05721) (to M. Hagiwara, and A.T.), Innovative Cell Biology by Innovating Technology (Cell Innovation) (to M. Hagiwara, K.O., and A.T.), a Core Research for Evolutional Science and Technology (CREST) grant from the Japan Science and Technology Agency (JST) (to M. Hagiwara), a grant from the Japan Agency for Medical Research and Development (AMED) (to M. Hagiwara), the Asian CORE Program of JSPS (to M. Hagiwara), iCeMS Cross-Disciplinary Research Promotion Project of Kyoto University (to A.T.), and the Fujiwara Memorial Foundation (to A.T.).

AUTHOR CONTRIBUTIONS

A.T. conceived and designed the project, performed the experiments, and wrote the manuscript. K.I. performed bioinformatic analysis. T.T. performed the Pol II ChIP-seq experiments. M.I. and K.O. performed the initial *in vivo* transcriptome experiment. M. Hosokawa assisted with the wet experiments. M.D. performed RNA-seq and ChIP-seq. K.N. performed the initial individual-nucleotide resolution CLIP (iCLIP) experiment. H. Kimura generated Pol II antibodies. T.A. and H. Kiyonari generated Sfpq conditional KO mice. J.B.B. assisted in writing the manuscript. M. Hagiwara conceived the project and prepared the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 1, 2017
Revised: January 19, 2018
Accepted: March 30, 2018
Published: May 1, 2018

REFERENCES

Allen, N.C., Bagade, S., McQueen, M.B., Ioannidis, J.P., Kavvoura, F.K., Khoury, M.J., Tanzi, R.E., and Bertram, L. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. Nat. Genet. *40*, 827–834.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Banerjee-Basu, S., and Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. Dis. Model. Mech. *3*, 133–135.

Bartkowiak, B., Liu, P., Phatnani, H.P., Fuda, N.J., Cooper, J.J., Price, D.H., Adelman, K., Lis, J.T., and Greenleaf, A.L. (2010). CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. Genes Dev. *24*, 2303–2316.

Bentley, D.L. (2014). Coupling mRNA processing with transcription in time and space. Nat. Rev. Genet. *15*, 163–175.

Berg, J.M., and Geschwind, D.H. (2012). Autism genetics: searching for specificity and convergence. Genome Biol. *13*, 247.

Blecher-Gonen, R., Barnett-Itzhaki, Z., Jaitin, D., Amann-Zalcenstein, D., Lara-Astiaso, D., and Amit, I. (2013). High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. Nat. Protoc. *8*, 539–554.

Bottini, S., Hamouda-Tekaya, N., Mategot, R., Zaragosi, L.E., Audebert, S., Pisano, S., Grandjean, V., Mauduit, C., Benahmed, M., Barbry, P., et al. (2017). Post-transcriptional gene silencing mediated by microRNAs is controlled by nucleoplasmic Sfpq. Nat. Commun. *8*, 1189.

Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat. Rev. Neurosci. *16*, 551–563.

Bradnam, K.R., and Korf, I. (2008). Longer first introns are a general property of eukaryotic gene structure. PLoS ONE *3*, e3093.

Buxadé, M., Morrice, N., Krebs, D.L., and Proud, C.G. (2008). The PSF.p54nrb complex is a novel Mnk substrate that binds the mRNA for tumor necrosis factor alpha. J. Biol. Chem. *283*, 57–65.

Chanas-Sacré, G., Mazy-Servais, C., Wattiez, R., Pirard, S., Rogister, B., Patton, J.G., Belachew, S., Malgrange, B., Moonen, G., and Leprince, P. (1999). Identification of PSF, the polypyrimidine tract-binding protein-associated splicing factor, as a developmentally regulated neuronal protein. J. Neurosci. Res. *57*, 62–73.

Chang, J., Gilman, S.R., Chiang, A.H., Sanders, S.J., and Vitkup, D. (2015). Genotype to phenotype relationships in autism spectrum disorders. Nat. Neurosci. *18*, 191–198.

Chathoth, K.T., Barrass, J.D., Webb, S., and Beggs, J.D. (2014). A splicing-dependent transcriptional checkpoint associated with prespliceosome formation. Mol. Cell *53*, 779–790.

Chen, L.L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol. Cell *35*, 467–478.

Cho, S., Moon, H., Loh, T.J., Oh, H.K., Williams, D.R., Liao, D.J., Zhou, J., Green, M.R., Zheng, X., and Shen, H. (2014). PSF contacts exon 7 of SMN2 pre-mRNA to promote exon 7 inclusion. Biochim. Biophys. Acta *1839*, 517–525.

Cortese, A., Plagnol, V., Brady, S., Simone, R., Lashley, T., Acevedo-Arozena, A., de Silva, R., Greensmith, L., Holton, J., Hanna, M.G., et al. (2014).

Widespread RNA metabolism impairment in sporadic inclusion body myositis TDP43-proteinopathy. Neurobiol. Aging *35*, 1491–1498.

Cosker, K.E., Fenstermacher, S.J., Pazyra-Murphy, M.F., Elliott, H.L., and Segal, R.A. (2016). The RNA-binding protein SFPQ orchestrates an RNA regulon to promote axon viability. Nat. Neurosci. *19*, 690–696.

Egloff, S., and Murphy, S. (2008). Cracking the RNA polymerase II CTD code. Trends Genet. *24*, 280–288.

Emili, A., Shales, M., McCracken, S., Xie, W., Tucker, P.W., Kobayashi, R., Blencowe, B.J., and Ingles, C.J. (2002). Splicing and transcription-associated proteins PSF and p54nrb/nonO bind to the RNA polymerase II CTD. RNA *8*, 1102–1111.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929–933.

Gabel, H.W., Kinde, B., Stroud, H., Gilbert, C.S., Harmin, D.A., Kastan, N.R., Hemberg, M., Ebert, D.H., and Greenberg, M.E. (2015). Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature *522*, 89–93.

Geschwind, D.H. (2011). Genetics of autism spectrum disorders. Trends Cogn. Sci. *15*, 409–416.

Hsin, J.P., and Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. Genes Dev. *26*, 2119–2137.

Iacobazzi, V., Infantino, V., Costanzo, P., Izzo, P., and Palmieri, F. (2005). Functional analysis of the promoter of the mitochondrial phosphate carrier human gene: identification of activator and repressor elements and their transcription factors. Biochem. J. *391*, 613–621.

Imamura, K., Imamachi, N., Akizuki, G., Kumakura, M., Kawaguchi, A., Nagata, K., Kato, A., Kawaguchi, Y., Sato, H., Yoneda, M., et al. (2014). Long noncoding RNA NEAT1-dependent SFPQ relocation from promoter region to paraspeckle mediates IL8 expression upon immune stimuli. Mol. Cell *53*, 393–406.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.H., Narzisi, G., Leotta, A., et al. (2012). De novo gene disruptions in children on the autistic spectrum. Neuron *74*, 285–299.

Ishigaki, S., Fujioka, Y., Okada, Y., Riku, Y., Udagawa, T., Honda, D., Yokoi, S., Endo, K., Ikenaka, K., Takagi, S., et al. (2017). Altered Tau Isoform Ratio Caused by Loss of FUS and SFPQ Function Leads to FTLD-like Phenotypes. Cell Rep. *18*, 1118–1131.

Ji, X., Zhou, Y., Pandit, S., Huang, J., Li, H., Lin, C.Y., Xiao, R., Burge, C.B., and Fu, X.D. (2013). SR proteins collaborate with 7SK and promoter-associated nascent RNA to release paused polymerase. Cell *153*, 855–868.

Kameoka, S., Duque, P., and Konarska, M.M. (2004). p54(nrb) associates with the 5′ splice site within large transcription/splicing complexes. EMBO J. *23*, 1782–1791.

Kaneko, S., Rozenblatt-Rosen, O., Meyerson, M., and Manley, J.L. (2007). The multifunctional protein p54nrb/PSF recruits the exonuclease XRN2 to facilitate pre-mRNA 3′ processing and transcription termination. Genes Dev. *21*, 1779–1789.

Ke, Y.D., Dramiga, J., Schütz, U., Kril, J.J., Ittner, L.M., Schröder, H., and Götz, J. (2012). Tau-mediated nuclear depletion and cytoplasmic accumulation of SFPQ in Alzheimer's and Pick's disease. PLoS ONE *7*, e35678.

Keene, J.D. (2007). RNA regulons: coordination of post-transcriptional events. Nat. Rev. Genet. *8*, 533–543.

Kimura, H., Hayashi-Takanaka, Y., Goto, Y., Takizawa, N., and Nozaki, N. (2008). The organization of histone H3 modifications as revealed by a panel of specific monoclonal antibodies. Cell Struct. Funct. *33*, 61–73.

King, I.F., Yandava, C.N., Mabb, A.M., Hsiao, J.S., Huang, H.S., Pearson, B.L., Calabrese, J.M., Starmer, J., Parker, J.S., Magnuson, T., et al. (2013). Topoisomerases facilitate transcription of long genes linked to autism. Nature *501*, 58–62.

Koga, M., Satoh, T., Takasaki, I., Kawamura, Y., Yoshida, M., and Kaida, D. (2014). U2 snRNP is required for expression of the 3′ end of genes. PLoS ONE *9*, e98015.

Koga, M., Hayashi, M., and Kaida, D. (2015). Splicing inhibition decreases phosphorylation level of Ser2 in Pol II CTD. Nucleic Acids Res. *43*, 8258–8267.

Lagier-Tourenne, C., Polymenidou, M., Hutt, K.R., Vu, A.Q., Baughn, M., Huelga, S.C., Clutario, K.M., Ling, S.C., Liang, T.Y., Mazur, C., et al. (2012). Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. Nat. Neurosci. *15*, 1488–1497.

Lee, M., Sadowska, A., Bekere, I., Ho, D., Gully, B.S., Lu, Y., Iyer, K.S., Trewhella, J., Fox, A.H., and Bond, C.S. (2015). The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. Nucleic Acids Res. *43*, 3826–3840.

Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron *70*, 886–897.

Licatalosi, D.D., Mele, A., Fak, J.J., Ule, J., Kayikci, M., Chi, S.W., Clark, T.A., Schweitzer, A.C., Blume, J.E., Wang, X., et al. (2008). HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature *456*, 464–469.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Lowery, L.A., Rubin, J., and Sive, H. (2007). Whitesnake/sfpq is required for cell survival and neuronal development in the zebrafish. Dev. Dyn. *236*, 1347–1357.

Lutz, C.S., Cooke, C., O'Connor, J.P., Kobayashi, R., and Alwine, J.C. (1998). The snRNP-free U1A (SF-A) complex(es): identification of the largest subunit as PSF, the polypyrimidine-tract binding protein-associated splicing factor. RNA *4*, 1493–1499.

Marshall, N.F., Peng, J., Xie, Z., and Price, D.H. (1996). Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase. J. Biol. Chem. *271*, 27176–27183.

Muñoz, M.J., de la Mata, M., and Kornblihtt, A.R. (2010). The carboxy terminal domain of RNA polymerase II and alternative splicing. Trends Biochem. Sci. *35*, 497–504.

O'Brien, T., Hardin, S., Greenleaf, A., and Lis, J.T. (1994). Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. Nature *370*, 75–77.

O'Connor, J.P., Alwine, J.C., and Lutz, C.S. (1997). Identification of a novel, non-snRNP protein complex containing U1A protein. RNA *3*, 1444–1455.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012). Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

Odawara, J., Harada, A., Yoshimi, T., Maehara, K., Tachibana, T., Okada, S., Akashi, K., and Ohkawa, Y. (2011). The classification of mRNA expression levels by the phosphorylation state of RNAPII CTD based on a combined genome-wide approach. BMC Genomics *12*, 516.

Oh, J.M., Di, C., Venters, C.C., Guo, J., Arai, C., So, B.R., Pinto, A.M., Zhang, Z., Wan, L., Younis, I., and Dreyfuss, G. (2017). U1 snRNP telescripting regulates a size-function-stratified human genome. Nat. Struct. Mol. Biol. *24*, 993–999.

Patton, J.G., Porro, E.B., Galceran, J., Tempst, P., and Nadal-Ginard, B. (1993). Cloning and characterization of PSF, a novel pre-mRNA splicing factor. Genes Dev. *7*, 393–406.

Peng, R., Dye, B.T., Pérez, I., Barnard, D.C., Thompson, A.B., and Patton, J.G. (2002). PSF and p54nrb bind a conserved stem in U5 snRNA. RNA *8*, 1334–1347.

Polymenidou, M., Lagier-Tourenne, C., Hutt, K.R., Huelga, S.C., Moran, J., Liang, T.Y., Ling, S.C., Sun, E., Wancewicz, E., Mazur, C., et al. (2011). Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat. Neurosci. *14*, 459–468.

Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., Gershoni, M., Morrey, C.P., Safran, M., and Lancet, D. (2017). MalaCards: an

amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. Nucleic Acids Res. *45* (*D1*), D877–D887.

Ray, P., Kar, A., Fushimi, K., Havlioglu, N., Chen, X., and Wu, J.Y. (2011). PSF suppresses tau exon 10 inclusion by interacting with a stem-loop structure downstream of exon 10. J. Mol. Neurosci. *45*, 453–466.

Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., Li, X., Gueroussov, S., Albu, M., Zheng, H., Yang, A., et al. (2013). A compendium of RNA-binding motifs for decoding gene regulation. Nature *499*, 172–177.

Rogelj, B., Easton, L.E., Bogu, G.K., Stanton, L.W., Rot, G., Curk, T., Zupan, B., Sugimoto, Y., Modic, M., Haberman, N., et al. (2012). Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. Sci. Rep. *2*, 603.

Rosonina, E., Ip, J.Y., Calarco, J.A., Bakowski, M.A., Emili, A., McCracken, S., Tucker, P., Ingles, C.J., and Blencowe, B.J. (2005). Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo. Mol. Cell. Biol. *25*, 6734–6746.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., et al. (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature *485*, 237–241.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat. Rev. Mol. Cell Biol. *7*, 557–567.

Singh, R.K., and Cooper, T.A. (2012). Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. *18*, 472–482.

Song, X., Sui, A., and Garen, A. (2004). Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: effects on steroidogenesis and oncogenesis. Proc. Natl. Acad. Sci. USA *101*, 621–626.

Stasevich, T.J., Hayashi-Takanaka, Y., Sato, Y., Maehara, K., Ohkawa, Y., Sakata-Sogawa, K., Tokunaga, M., Nagase, T., Nozaki, N., McNally, J.G., and Kimura, H. (2014). Regulation of RNA polymerase II activation by histone acetylation in single living cells. Nature *516*, 272–275.

Sun, J., Jia, P., Fanous, A.H., Webb, B.T., van den Oord, E.J., Chen, X., Bukszar, J., Kendler, K.S., and Zhao, Z. (2009). A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases-schizophrenia as a case. Bioinformatics *25*, 2595–6602.

Thomas-Jinu, S., Gordon, P.M., Fielding, T., Taylor, R., Smith, B.N., Snowden, V., Blanc, E., Vance, C., Topp, S., Wong, C.H., et al. (2017). Non-nuclear Pool of Splicing Factor SFPQ Regulates Axonal Transcripts Required for Normal Motor Development. Neuron *94*, 931.

Ule, J., Jensen, K.B., Ruggiu, M., Mele, A., Ule, A., and Darnell, R.B. (2003). CLIP identifies Nova-regulated RNA networks in the brain. Science *302*, 1212–1215.

Urban, R.J., Bodenburg, Y., Kurosky, A., Wood, T.G., and Gasic, S. (2000). Polypyrimidine tract-binding protein-associated splicing factor is a negative regulator of transcriptional activity of the porcine p450scc insulin-like growth factor response element. Mol. Endocrinol. *14*, 774–782.

Van Nostrand, E.L., Pratt, G.A., Shishkin, A.A., Gelboin-Burkhart, C., Fang, M.Y., Sundararaman, B., Blue, S.M., Nguyen, T.B., Surka, C., Elkins, K., et al. (2016). Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). Nat. Methods *13*, 508–514.

Veltman, J.A., and Brunner, H.G. (2012). De novo mutations in human genetic disease. Nat. Rev. Genet. *13*, 565–575.

Yang, J., Zhao, Y., Kalita, M., Li, X., Jamaluddin, M., Tian, B., Edeh, C.B., Wiktorowicz, J.E., Kudlicki, A., and Brasier, A.R. (2015). Systematic Determination of Human Cyclin Dependent Kinase (CDK)-9 Interactome Identifies Novel Functions in RNA Splicing Mediated by the DEAD Box (DDX)-5/17 RNA Helicases. Mol. Cell. Proteomics *14*, 2701–2721.

Yarosh, C.A., Iacona, J.R., Lutz, C.S., and Lynch, K.W. (2015). PSF: nuclear busy-body or nuclear facilitator? Wiley Interdiscip. Rev. RNA *6*, 351–367.

Zhou, Q., Li, T., and Price, D.H. (2012). RNA polymerase II elongation control. Annu. Rev. Biochem. *81*, 119–143.